

## The publication of press releases as journalistic information. Comparative study of two Spanish newspapers

Jesús Martínez Fernández<sup>1</sup>; Francesc Pozo Montero<sup>2</sup>; Josep Lluís Micó Sanz<sup>3</sup>

Recibido: 13 de marzo de 2019 / Aceptado: 24 de junio de 2019

**Abstract.** The distinction of what constitutes a “news event” can give rise to many interpretations. In this world of telematic accessibility, which is a consequence of globalization, events and occurrences of all kinds can be categorized as news simply by dressing them up as news. According to the style and communications manuals, the news has its own characteristics: relevance, social interest and proximity, among others. Press releases have become perfected as a result of increasingly sophisticated public relations agencies, and with them the thin line between information and advertising is now blurred. In this paper, we compare press releases issued by public and private companies with news briefs in the Economy sections of newspapers. As will be seen, many of them coincide and bear some similarities. The sample uses news briefs published during the first half of 2014 in *El Mundo* and *La Vanguardia*, Spanish pay-to-read newspapers that feature prominently in the analysis by the Estudio General de Medios (General Media Study). The methodology makes use of the Maple program with its DetectPlagiarism command to perform an *ad hoc* comparison of the texts. The default copy threshold for DetectPlagiarism is 0.35. The similarity indices between the news briefs and the press releases of *La Vanguardia* and *El Mundo* indicate values higher than this threshold.

**Keywords:** Journalism; genres; plagiarism; economy; companies; business communication.

### [es] La publicación de comunicados de prensa como información periodística. Estudio comparativo de dos diarios españoles

**Resumen.** La distinción de lo que constituye un “evento noticiable” puede dar lugar a muchas interpretaciones. En este mundo de accesibilidad telemática, que es una consecuencia de la globalización, los eventos y los sucesos de todo tipo se pueden clasificar como noticias simplemente vistiéndolos como noticias. De acuerdo con los manuales de estilo y comunicación, la noticia tiene características propias: relevancia, interés social y proximidad, entre otras. Los comunicados de prensa se han perfeccionado como resultado de las agencias de relaciones públicas cada vez más sofisticadas, y con ellas la línea delgada entre la información y la publicidad ahora está borrosa. En este artículo, comparamos comunicados de prensa emitidos por empresas públicas y privadas con breves publicados en las secciones de economía de los periódicos. Como se verá, muchos de ellos coinciden y tienen algunas similitudes. La muestra utiliza breves publicados durante el primer semestre de 2014 en *El Mundo* y *La Vanguardia*, los periódicos en español de pago por lectura que ocupan un lugar destacado en el análisis del Estudio General de Medios. La metodología hace uso del programa Maple con su comando DetectPlagiarism para realizar una comparación *ad hoc* de los textos. El umbral de copia predeterminado para DetectPlagiarism es 0.35. Los índices de similitud entre los breves y los comunicados de prensa de *La Vanguardia* y *El Mundo* indican valores superiores a este umbral.

**Palabras clave:** Periodismo; géneros; plagio; economía; empresas; comunicación empresarial.

**Sumario.** 1. Introduction 2. Objectives and methodology 3. Results 3.1. La Vanguardia 3.2 El Mundo 4. Conclusions 5. References

**Cómo citar:** Martínez Fernández, Jesús; Pozo Montero, Francesc; Micó Sanz, Josep Lluís (2020): “The publication of press releases as a journalistic information. Comparative study of two Spanish newspapers”. *Estudios sobre el Mensaje Periodístico* 26(1), 215-228.

### 1. Introducción

A group of journalism students under the guidance of reporter Wendy Bacon analysed the content of a dozen Australian newspapers for a week, and they concluded that more than half of the news published (55%) had originated in a press release. Their re-

search, sponsored by the Australian Center for Independent Journalism and the Crikey portal, led them to discover that there were staff journalists who signed pieces that were practically literal reproductions of information sent to them by companies and institutions. Of the 2,203 articles analysed, more than 500 contained no additional points of view,

<sup>1</sup> Universidad Autònoma de Barcelona (España)  
E-mail: [reporterojesus@gmail.com](mailto:reporterojesus@gmail.com)

<sup>2</sup> Universidad Politècnica de Catalunya (España)  
E-mail: [francesc.pozo@upc.edu](mailto:francesc.pozo@upc.edu)

<sup>3</sup> Universidad Ramon Llull (España)  
E-mail: [joseplluisms@blanquerna.url.edu](mailto:joseplluisms@blanquerna.url.edu)

sources or content beyond what was provided by the press release (Bacon, 2010).

This issue has become a major concern for media researchers, especially in precarious markets such as Spain, where editors are required to produce much while relying on scant resources and personnel.

Drawing on the high degree of “manipulation” that occurs in the informational chain (Durandin, 1982) as well as the digital chain (Scholz, 2013), other authors have made useful contributions about the use and abuse of agency information in a venue that is underscored by urgency and scarcity of means: the electronic editions of the newspapers.

It is for this reason that some authors refer to the misinformation that is generated in this way (Starr, 2004, 2009), because it is their opinion that what is really disseminated by media that operate on all platforms is advertising, propaganda and promotional content, which is ultimately not information that has been corroborated and prepared rigorously (Schwoebel, 1971, McChesney, 1999, Manning, 2001). However, no works to date have focused exclusively on this widespread practice in regard to the news brief, a genre that is frequently subject to this practice, as will be seen below. This line of research was not even pursued by researchers who have suggested it in their own papers (De Fontcuberta, 2000: 16, Parrott, 2008).

The news brief is not listed in the *Teoría de los géneros periodísticos* [Theory of the journalistic genres], by Llorenç Gomis (2008). However, it obviously exists. It is often presented as the antithesis of reporting, the “all-embracing” genre, the quintessential star (García Márquez, 2010). Nevertheless, news briefs fulfil a function as important as any other format. Authors such as Grijelmo (2006) highlight that it is a news item that barely consists of a lede, that is, the “essential core of the news”. Therefore, it can be understood not as a summary, but as an outline of an informative event that is communicated with great economy of words. Its function is to provide an immediate and clear account of what is fundamental while generating interest and attracting attention so that the public can continue to read, watch and listen (Jaraba, 2009).

Generally, the news brief covers two of the 6 W’s (*who, what, when, where, why* and *how*)—namely, who and what. But this is not a fixed norm: sometimes it is necessary that it attends to other W’s, especially if it is the follow-up to a news item that has been gaining traction in previous days. In any event, it can be written neatly and elegantly. For this reason, Grijelmo asks that the editors and copywriters be attentive to the news briefs that often appear in the papers so that all doubts about the news can be cleared up. In any case, the information in the brief—which is not the same as brief information—should follow the so-called informative style, which is characterized by being “sober and concise, objective, where there is no place for the journalist’s ego” (De Fontcuberta, 2000).

The newspaper stylebooks studied in this research treat the news brief in the following way. *La Vanguardia* (2004) reduces it to a subsection of the news genre (the others are the report, the interview, the chronicle, the analysis, the review and the op-ed):

The news briefs (or *newsflashes*) are pieces that gather news in short form. Grouped into a block of one or two columns, news briefs (minimum of two) have a similar extension and must contain all the minimum elements for understanding the information. They have no line breaks. These pieces are not dated, but authorship must be indicated at the end, separated from the text by a long dash between spaces: if it is from an editor, it is signed with initials or with an initial and surname (never with the whole name), or with the word “Editorial”; if it is an agency, the corresponding agency is indicated (*La Vanguardia*, 2004: 34).

The news brief is not listed independently in the *El Mundo* (2002) style book, but it does explain the content of what the “basic” news should be:

The news or basic information. This was the most common genre in the daily press, which applied more strict considerations to phrases and short paragraphs, a straight opening and developing the story in a way that can be chronological or pyramidal, depending on the greater or lesser complexity of the informative elements (*El Mundo*, 2002: 6).

## 2. Objectives and methodology

The objective of this research is to compare the press releases written by companies and public entities with news briefs about these organizations published in the Economy sections of the media included in this study (*El Mundo* and *La Vanguardia*). This comparison is intended to detect whether the press release is processed and corroborated before being published or if, on the contrary, the press release is distributed in the same form as it arrives in the newsroom. In this way we seek to establish numerically what percentage of the published news brief coincides with or is similar to the press release.

Although our efforts are not intended to accuse newspapers or news agencies of plagiarism, the tools that will be used have indeed been conceived for detecting plagiarism. In recent years, a large number of programs and applications have been created for this purpose, which intends to reduce and avoid copying in publications such as scientific journals (Sánchez-Vega, Villatoro-Tello, Montes-y-Gómez, Villaseñor–Pineda and Rosso, 2013).

One of the most widespread tools is iThenticate, whose comparative database contains billions of articles that the editors of more than five hundred media sources use to detect plagiarism and authorship problems.

Bibliotècnica (<https://bibliotecnica.upc.edu>), the digital library of the Universitat Politècnica de Cata-

lunya – BarcelonaTech, has drawn up an extensive list of programs that can be found on the internet for detecting plagiarism. In addition, there are programs that are content verification tools for journalists, such as Plagtracker, which is recommended by the International Center for Journalists.

In this environment, it is advisable to be familiar with the *Maple* computer algebra system program that is owned by Maplesoft. It is also helpful to understand the *command* concept in computer programs and applications, as well as each of the functions and orders whose purposes are to perform specific tasks.

More specifically in regard to the framework of this investigation's methodology, some of the *Maple* commands used are described below:

- **SimilarityScore:** Compares the use of common words between two texts and it returns a value or score between 0 and 1. A score of 0 indicates that there is no common word between both texts; a score of 1 indicates that there is a total match of words.
- **JaccardCoefficient:** Compares two texts. The result is a number between 0 and 1. The **CosineCoefficient** and **DiceCoefficient** commands are similar.
- **DetectPlagiarism:** Establishes whether plagiarism exists within two texts.

The sample covered the first half of 2014. *El Mundo* and *La Vanguardia* were chosen because these general information pay-to-read papers in Spain have more business news briefs as a fixed part of a section (Economy). In terms of the number of daily readers for this category, they occupy the second and third positions, respectively, behind *El País*—which was discarded because it lacked both the volume and the desired regularity in publishing news briefs in their Economy section (AIMC, 2017: 8). In some cases, the same news brief appeared in both newspapers on the same day. However, our priority lied not in whether news briefs coincided in the newspapers being compared, but rather to analyse news briefs sourced from a company press release, that is, briefs that originated with a commercial enterprise and were sent to the paper for promotional purposes.

News briefs were chosen at random in order to increase the probability of arriving at findings more likely to be representative of a group (Csikszentmihalyi and Larson, 2014: 35-54). The selection of news briefs followed the following process. We began with the Wednesday, January 1, 2014 editions of the two newspapers. This was followed by taking a selection from the next week day, Thursday, but for the following week, i.e., Thursday, January 9, 2014. Thus, in a scaled fashion, we covered half a year for the two dailies.

In most cases, the authorship of the news brief was an agency. Europa Press is the agency preferred by *La Vanguardia* and *El Mundo* in their respective Economy sections, more so than EFE. They also use the signature “Agencies” (without specifying) or “Editorial Staff” (or, in the case of *El Mundo*, E. M.)

As explained previously, there are many applications for detecting plagiarism, and research in this field is extensive and growing (Barrón-Cedeño, Gupta and Rosso, 2013). This investigation compares the news brief with the press release or the agency dispatch—which can easily be located on the internet. Thus, the problem was less complex than what is generally the case with scientific literature.

Based on these considerations, we chose *Maple's* EssayTools library for detecting plagiarism in the field of journalism. This tool was chosen because it met the needs of this study—namely, comparing two texts—and because of the programming's flexibility in the *Maple* environment.

In this context, the **SimilarityScore** command is used to compare the use of words in two or more texts, and it returns a matrix of scores for each pair of texts. A score of 0 indicates that there is no overlap between the two. Conversely, a score of 1 indicates that there is a total overlap. However, a score of 1 does not necessarily imply that both texts are identical.

For example:

```
> with (EssayTools);
> textA: = “Above the beautiful rain fell”;
> textB: = “Above fell the beautiful rain”;
> SimilarityScore (textA, textB)
[1]
```

While:

```
> with (EssayTools);
> textC: = “Paris Rome”;
> textD: = “London Barcelona”;
> SimilarityScore (textC, textD)
[0]
```

There are also several methods available to compare the level of similarity between two texts, such as the so-called **CosineCoefficient**, **JaccardCoefficient**, **DiceCoefficient** and their binary equivalents (Niwatanakul, Singthongchai, Naenudorn and Wanapu, 2013).

If for example the texts “The tall day” and “The tall dog” are considered and a list is made with all the different words that appear in both texts, the list is [the, day, tall, dog]. Next, two vectors are generated that indicate how many times these words appear in the texts. The vector for the first text is [1 1 1 0], since “the” appears once, as well as “day” and “tall”, while “dog” does not appear. The vector for the second text is [1 0 1 1]. These two vectors will be used to perform a mathematical operation, which will be different depending on the method used. In the case of **Cosine Coefficient** and given two vectors  $v_1$  and  $v_2$ , the cosine coefficient is calculated as the scalar product of both vectors between the product of their norms, that is:

$$\frac{v_1 \cdot v_2}{|v_1| \cdot |v_2|}$$

where:

$$v_1 \cdot v_2 = \sum_{i=1}^n v_1(i)v_2(i) \in \mathbb{R}$$

$$|v_1| = \sqrt{\sum_{i=1}^n v_1(i)^2} \geq 0$$

$$|v_2| = \sqrt{\sum_{i=1}^n v_2(i)^2} \geq 0$$

For the vectors that were used as an example, we have:

$$v_1 = [ 1 \ 1 \ 1 \ 0 ]$$

$$v_2 = [ 1 \ 0 \ 1 \ 1 ]$$

$$v_1 \cdot v_2 = 2$$

$$|v_1| = \sqrt{3}$$

$$|v_2| = \sqrt{3}$$

Thus:

$$\frac{v_1 \cdot v_2}{|v_1| \cdot |v_2|} = \frac{2}{3}$$

Each of these three coefficients has a binary version in which the number of times a word appears is not important, but whether or not it does appear is indeed important.

DetectPlagiarism compares texts through the SimilarityScore command and returns those that exceed a certain similarity threshold. It returns the

texts that are so similar that they are likely to have been copied or they contain some portion of the text that has been copied. This is a probabilistic measure and does not necessarily imply that two texts with a high score are indeed copies of each other or that they have been copied from the same source. The default threshold in the DetectPlagiarism command is 0.35 and the similarity metric is BinaryJaccard-Coefficient. Both the threshold and similarity metric can be modified to reach any value between 0 and 1 (Wang, Qi, Kong and Nu, 2013).

When comparing two texts, the length of both is important. More precisely, the result of the comparison can be altered if a very short text is compared with another very long text. To solve this problem while keeping in mind that for the present investigation the shortest text is always the news brief, a recurrent comparison was made between the news brief and an increasingly large part of the press release or the wire service dispatch. This was carried out through an algorithm whose result is the maximum value of the similarity index and the number of words that maximize this similarity to the press release or wire dispatch.

Using this methodology, the statistical study was carried out on a sample of 52 news briefs from the newspapers *La Vanguardia* and *El Mundo*: 30 pieces by *La Vanguardia* and 22 pieces by *El Mundo*. Henceforth, the abbreviations used correspond to *La Vanguardia* (LV), *El Mundo* (EM), press release (PR) and a dispatch or news release from a news agency (NA). A cross indicates the existence of a press release and/or an agency dispatch / news release (Table I).

Table I. The statistical study was carried out on a sample of 52 news briefs from the newspapers *La Vanguardia* (LV) and *El Mundo* (EM)

<b>LV</b>	<b>1</b>	<b>3</b>	<b>5</b>	<b>7</b>	<b>9</b>	<b>11</b>	<b>13</b>	<b>15</b>	<b>17</b>	<b>19</b>	<b>21</b>	<b>23</b>	<b>25</b>	<b>27</b>	<b>29</b>
<b>PR</b>	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
<b>AD</b>						X		X		X			X	X	X
	<b>31</b>	<b>33</b>	<b>35</b>	<b>37</b>	<b>39</b>	<b>41</b>	<b>43</b>	<b>45</b>	<b>46</b>	<b>47</b>	<b>48</b>	<b>49</b>	<b>50</b>	<b>51</b>	<b>52</b>
<b>PR</b>	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
<b>AD</b>			X			X	X	X	X	X			X	X	
<b>EM</b>	<b>2</b>	<b>4</b>	<b>6</b>	<b>8</b>	<b>10</b>	<b>12</b>	<b>14</b>	<b>16</b>	<b>18</b>	<b>20</b>	<b>22</b>				
<b>PR</b>	X	X	X	X	X	X	X	X	X	X	X				
<b>AD</b>			X			X	X	X	X	X					
	<b>31</b>	<b>33</b>	<b>35</b>	<b>37</b>	<b>39</b>	<b>41</b>	<b>43</b>	<b>45</b>	<b>46</b>	<b>47</b>	<b>48</b>				
<b>PR</b>	X	X	X	X	X	X	X	X	X	X	X				
<b>AD</b>			X			X	X	X	X	X					

After collecting the data (carried out in the first half of 2014), it was systematically ordered and analysed, work which lasted into the second half of 2014. To expand the scope of the research (Wolf, 1987; Wimmer and Dominick, 2013), a series of in-depth semistructured interviews were conducted. This was done throughout 2015 and the first quarter of 2016 in a first round, then a second round took

place in the first four months of 2018. The interpretation of the participants in this second phase was fundamental to understanding all the nuances of the numerical results obtained (Busquet, Medina and Sort, 2006), for which we relied on the following profile of interviewees: workers in the media being studied, both editors and department or section managers; employees and managers in press offices and com-

munications agencies; and journalism teachers and researchers with experience in this professional field. In total, twenty people were interviewed, equally divided between women and men with varying degrees of experience: at one end of the spectrum there was a corporate communications assistant with less than ten years of experience, and at the other a newspaper executive with several decades of experience.

### 3. Results

We separately compared news briefs with press releases; news briefs with agency dispatches; and press releases with either dispatches or news releases from agencies. Of the 30 news briefs analysed from *La Vanguardia*, 19 resemble the press releases of the multinationals. If an intermediary exists between the news brief and the company press release (i.e., a news agency), the similarity is also remarkable: in 14 of these 30 news briefs of *La Vanguardia* that rely on an agency dispatch, the agency dispatch coincides with the news brief in 10 cases. What is more, the dispatch coincides fully with the original press release. When the agency dispatch and the corresponding press release are available (14 cases), the news brief and the press release of the multinational are similar on 9 occasions.

As for the 22 news briefs analysed from *El Mundo*, 12 have aspects that coincide with the wording proposed by the multinational. In this newspaper,

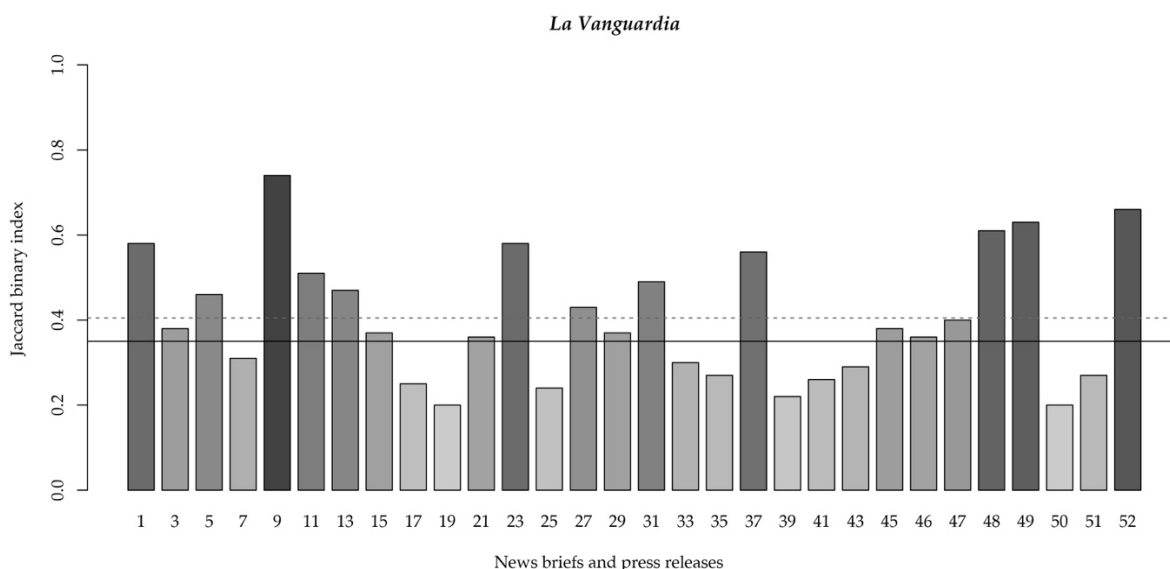
the intermediary (i.e., the news agency) intervenes most (this is the case in 21 of the 22 cases). Thus, the news brief maintains clear similarities to the agency piece. Of the 19 cases in which both the agency dispatch and the multinational’s original wording are available, homogeneity between them occurs 13 times.

#### 3.1. La Vanguardia

Using the similarity algorithm, we have compared the 30 news briefs of *La Vanguardia* with their corresponding press releases. The results appear in detail in Table II and are summarized graphically in Image 1. In Images 1-8, the solid line represents the threshold of similarity, above which it can be said that similarity does exist; the dashed line marks the arithmetic mean of the similarity indexes and the intensity of grey in each of the columns is proportional to the similarity coefficient. In the same way, Tables II–VII include the length of the text necessary for achieving maximum similarity and the final test result. When the Jaccard binary index is higher than 0.35, there is statistical evidence of great similarity. This is reflected in the last column of Table II.

Another way of looking at this is the following. Of the 30 pairs of texts that were compared, 19 of them obtained a similarity value higher than 0.35, which represents 63.33% of the texts. What is more, the average value of the similarity index is 0.41, which is higher than 0.35.

Image 1. Jaccard binary index when comparing news briefs with press releases



Source: Own elaboration

Table II. Jaccard binary index, result of comparing the indicated news briefs and the press releases

Text #1	Text #2	Index	Words	Similarity?
<b>B01</b>	PR01	0,58	75	Yes
<b>B03</b>	PR03	0,38	57	Yes

<b>B05</b>	PR05	0,46	133	Yes
<b>B07</b>	PR07	0,31	52	No
<b>B09</b>	PR09	0,74	93	Yes
<b>B11</b>	PR11	0,51	72	Yes
<b>B13</b>	PR13	0,47	57	Yes
<b>B15</b>	PR15	0,37	64	Yes
<b>B17</b>	PR17	0,25	94	No
<b>B19</b>	PR19	0,20	91	No
<b>B21</b>	PR21	0,36	114	Yes
<b>B23</b>	PR23	0,58	82	Yes
<b>B25</b>	PR25	0,24	105	No
<b>B27</b>	PR27	0,43	72	Yes
<b>B29</b>	PR29	0,37	81	Yes
<b>B31</b>	PR31	0,49	50	Yes
<b>B33</b>	PR33	0,30	97	No
<b>B35</b>	PR35	0,27	75	No
<b>B37</b>	PR37	0,56	77	Yes
<b>B39</b>	PR39	0,22	52	No
<b>B41</b>	PR41	0,26	55	No
<b>B43</b>	PR43	0,29	59	No
<b>B45</b>	PR45	0,38	57	Yes
<b>B46</b>	PR46	0,36	86	Yes
<b>B47</b>	PR47	0,40	85	Yes
<b>B48</b>	PR48	0,61	86	Yes
<b>B49</b>	PR49	0,63	86	Yes
<b>B50</b>	PR50	0,20	142	No
<b>B51</b>	PR51	0,27	104	No
<b>B52</b>	PR52	0,66	55	Yes

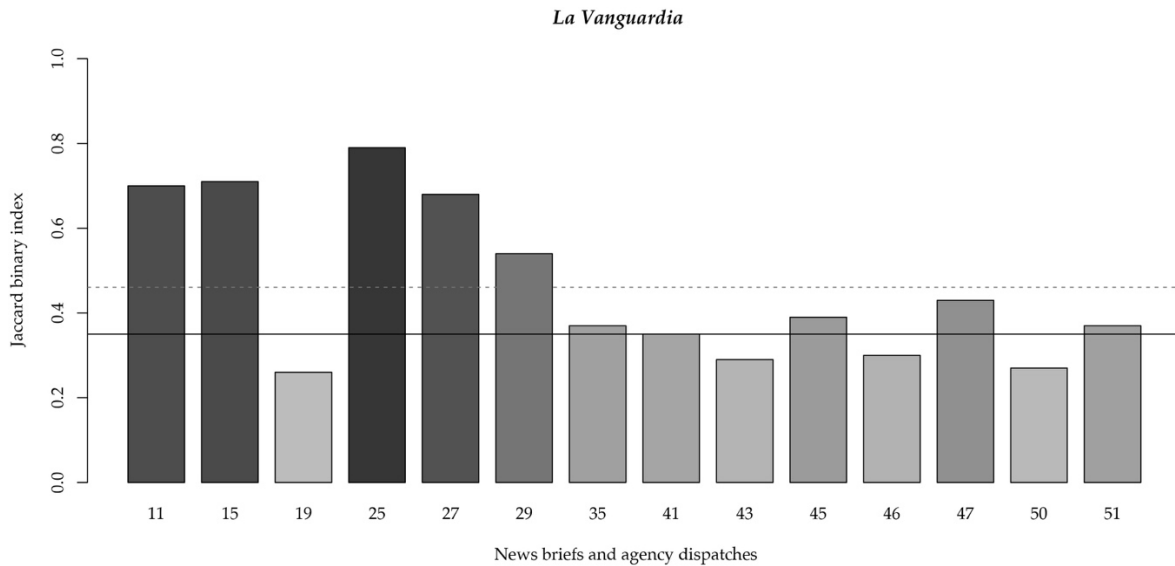
With the similarity algorithm, we have compared the *La Vanguardia* news briefs and the corresponding agency dispatches. The results, which are summarized in Image 2, appear in detail in Table III. It can be seen that of the 14 pairs of texts

compared, a similarity value greater than 0.35 is obtained in 10, which represents 71.43% of the texts. In addition, the average value of the similarity index stands at 0.46, which is higher than 0.35. Ten of the 14 news briefs coincide.

Table III. Jaccard binary index, result of comparing the indicated news briefs and the agency news releases

Text #1	Text #2	Index	Words	Similarity?
<b>B11</b>	AD11	0,70	63	Yes
<b>B15</b>	AD15	0,71	76	Yes
<b>B19</b>	AD19	0,26	88	No
<b>B25</b>	AD25	0,79	107	Yes
<b>B27</b>	AD27	0,68	96	Yes
<b>B29</b>	AD29	0,54	86	Yes
<b>B35</b>	AD35	0,37	81	Yes
<b>B41</b>	AD41	0,35	53	Yes
<b>B43</b>	AD43	0,29	69	No
<b>B45</b>	AD45	0,39	62	Yes
<b>B46</b>	AD46	0,30	128	No
<b>B47</b>	AD47	0,43	128	Yes
<b>B50</b>	AD50	0,27	144	No
<b>B51</b>	AD51	0,37	85	Yes

Image 2. Jaccard binary index when comparing news briefs with agency dispatches



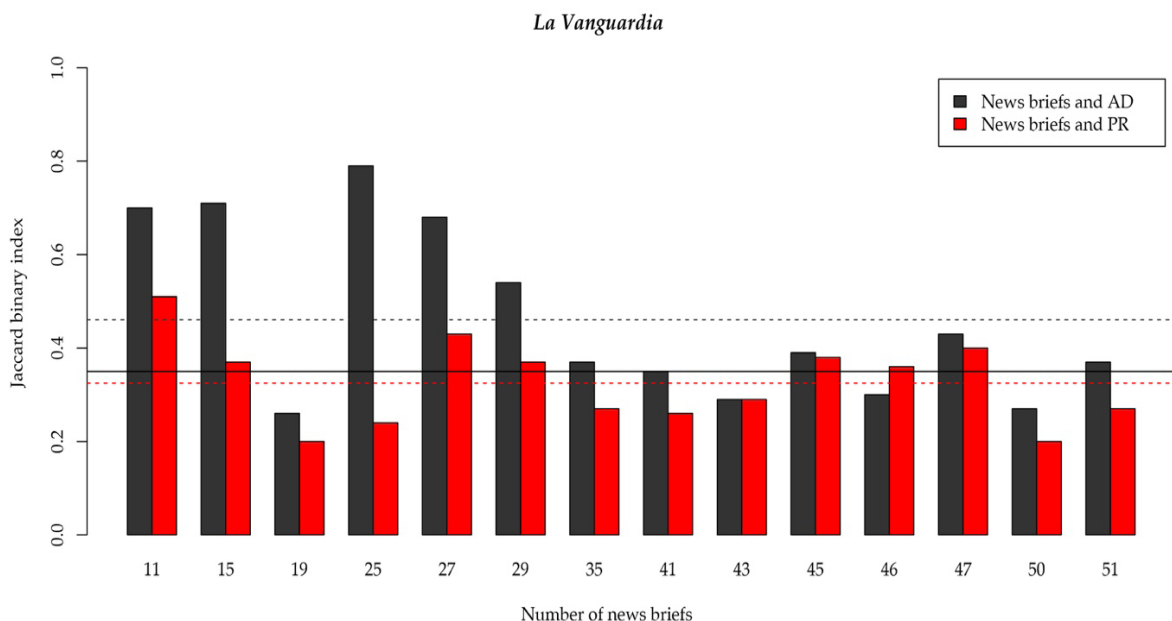
Source: Own elaboration

In those cases, in which both the press release and the agency dispatch are available, a comparison has also been established that is summarized graphically in Image 3. In 14 of the 30 *La Vanguardia* news briefs, the agency dispatch has been obtained as well as the original source, i. e., the company’s statement. In observing the abovementioned illustration, it is clear that in all but one case the similarity index between the news brief and agency dispatch is higher than the index between the news brief and the press release. The average value (0.33) of the similarity indices between the news brief and the press release—when the news agency dispatch is also available—is below the 0.35 threshold. In contrast, the average of

the similarity indices between news brief and agency dispatch is, in this case, 0.46.

For *La Vanguardia*, when the press release and agency news release are available, the byline given to the published news brief is, in most cases, a news agency (Europa Press, EFE or even “Agencies”). The similarity algorithm compares the similarity between the agency news releases that are the source of the news briefs and the original press releases. The results are summarized in Image 4 and are broken down in Table IV. It can be seen that, out of the 14 pairs of texts compared, a similarity value that is clearly higher than 0.35 is reached in 9 of them, representing 64.28% of the texts. The average value of the similarity index is 0.41, which is higher than 0.35.

Image 3 Jaccard binary index when comparing news briefs with agency dispatches (black columns) and with press releases (red columns)

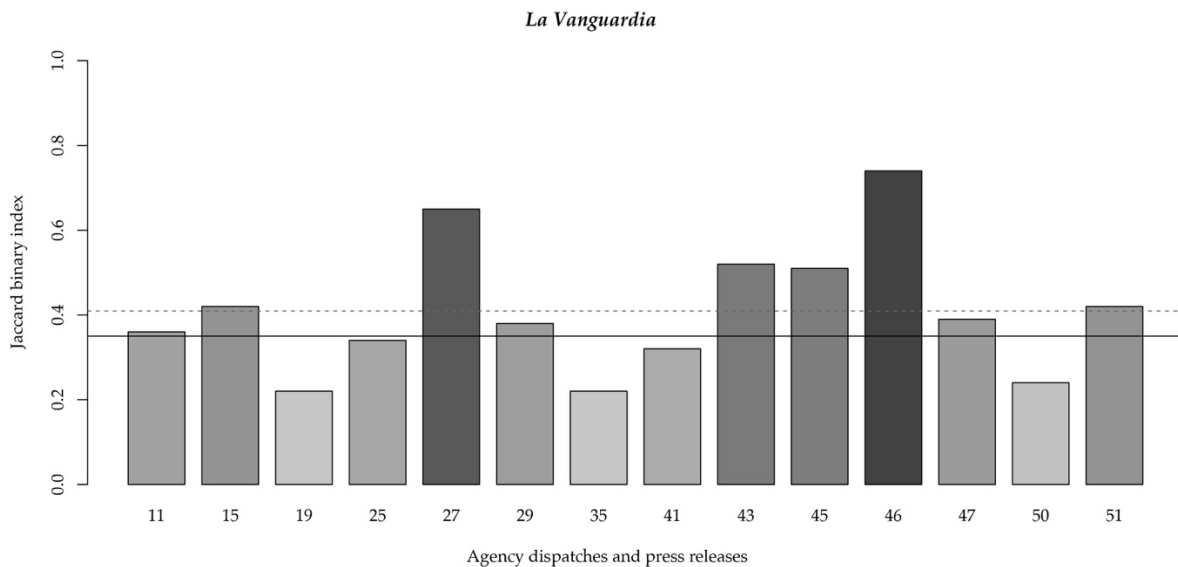


Source: Own elaboration

Table IV. Jaccard binary index, result of comparing the indicated press releases and the agency information

Text #1	Text #2	Index	Words	Similarity?
AD11	PR11	0,38	201	Yes
AD15	PR15	0,42	244	Yes
AD19	PR19	0,22	519	No
AD25	PR25	0,34	400	No
AD27	PR27	0,65	318	Yes
AD29	PR29	0,38	140	Yes
AD35	PR35	0,22	219	No
AD41	PR41	0,32	432	No
AD43	PR43	0,52	383	Yes
AD45	PR45	0,51	149	Yes
AD46	PR46	0,74	263	Yes
AD47	PR47	0,39	135	Yes
AD50	PR50	0,24	261	No
AD51	PR51	0,42	282	Yes

Image 4. Jaccard binary index when comparing press releases with agency dispatches



Source: Own elaboration

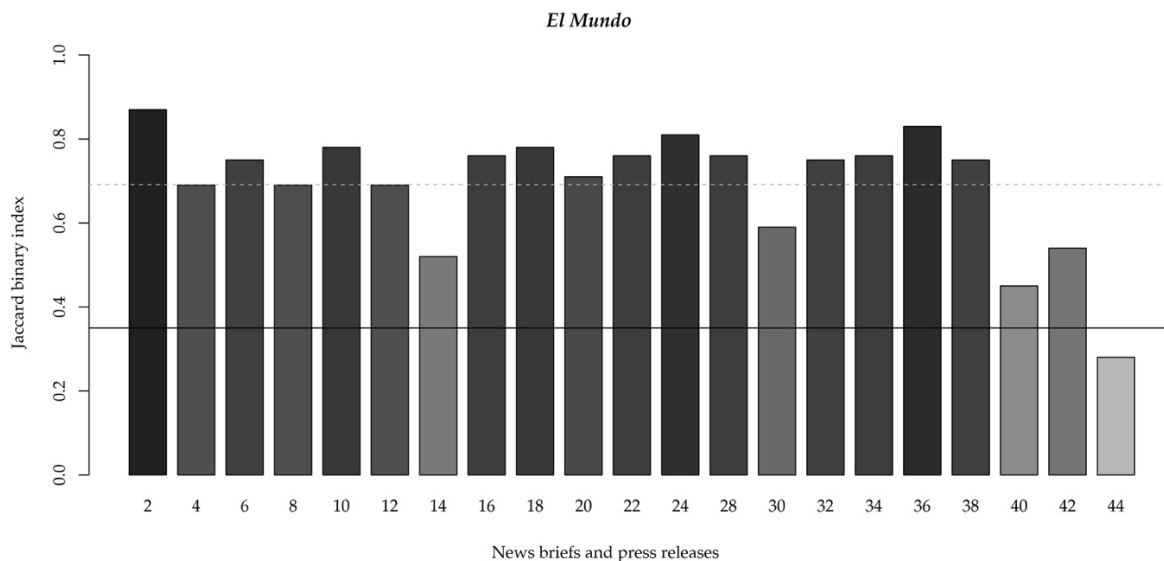
### 3.2. El Mundo

In the same way that we have proceeded with *La Vanguardia*, we have compared the similarity between *El Mundo's* news briefs and the corresponding press releases. The results are summarized in Image 5 and are more greatly detailed in Table V. When Jaccard's

binary index is greater than 0.35, statistical evidence of great similarity is indicated, which is reflected in the last column of Table V. In 12 of the 20 pairs of compared texts, a similarity value of more than 0.35 is reached, which represents 60% of the texts. The average value of the similarity index is 0.41, which is higher than 0.35.



Image 5. Jaccard binary index when comparing news briefs with press releases



Source: Own elaboration

Table V. Jaccard binary index, result of comparing the indicated news briefs and the press releases

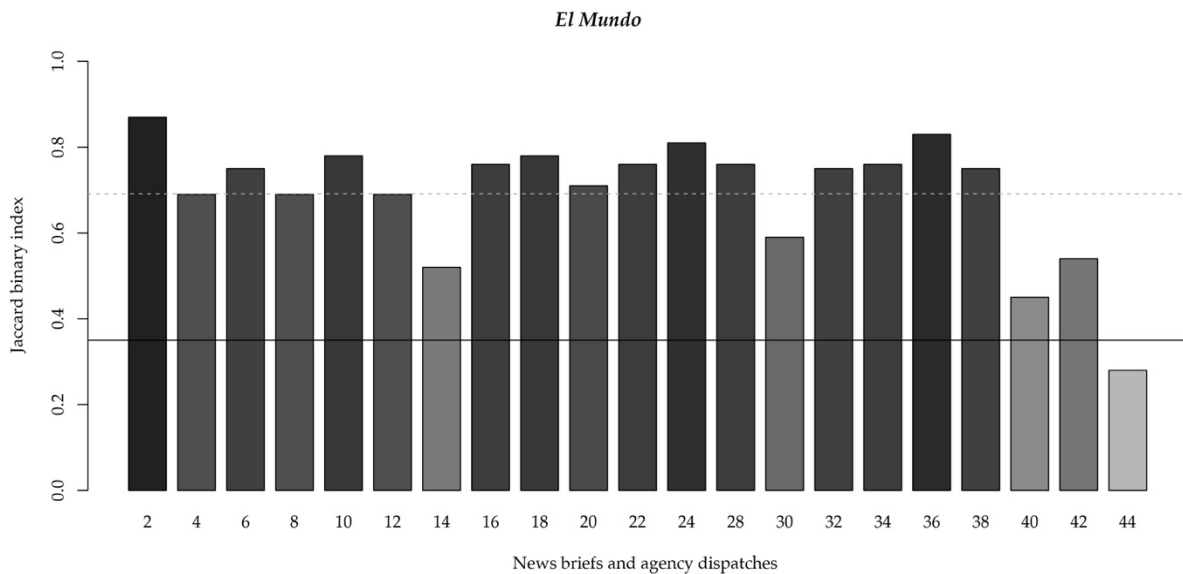
Text #1	Text #2	Index	Words	Similarity?
<b>B04</b>	PR04	0,25	139	No
<b>B06</b>	PR06	0,43	160	Yes
<b>B08</b>	PR08	0,55	102	Yes
<b>B10</b>	PR10	0,49	52	Yes
<b>B12</b>	PR12	0,41	90	Yes
<b>B14</b>	PR14	0,32	104	No
<b>B16</b>	PR16	0,23	59	No
<b>B16</b>	PR18	0,26	44	No
<b>B20</b>	PR20	0,16	36	No
<b>B22</b>	PR22	0,46	149	Yes
<b>B24</b>	PR24	0,35	99	Yes
<b>B26</b>	PR26	0,77	66	Yes
<b>B28</b>	PR28	0,23	127	No
<b>B30</b>	PR30	0,46	57	Yes
<b>B32</b>	PR32	0,72	77	Yes
<b>B34</b>	PR34	0,75	95	Yes
<b>B36</b>	PR36	0,21	48	No
<b>B40</b>	PR40	0,53	74	Yes
<b>B42</b>	PR42	0,36	61	Yes
<b>B44</b>	PR44	0,23	64	No

Source: Own elaboration

Likewise, the similarity algorithm has been used to compare the similarity between *El Mundo's* news briefs and the corresponding dispatches or agency news releases. The results are presented as above in Image 6 and Table VI. In 20 of the 21 matched pairs, a similarity value greater than 0.35

is obtained, which represents 95.24% of the texts. The average value of the similarity index stands at 0.69, which is also higher than 0.35. The similarity established between the news brief and the agency dispatch is greater than that between the news brief and the press release.

Image 6 Jaccard binary index when comparing the news briefs with the agency news releases



Source: Own elaboration

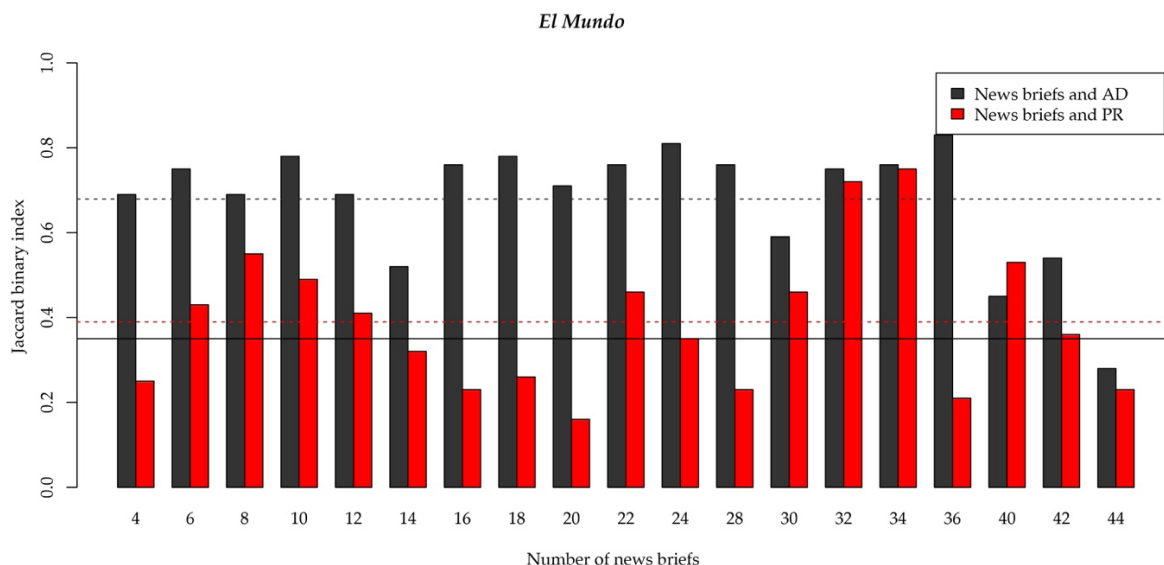
Table VI. Jaccard binary index, result of comparing the indicated news briefs and the agency dispatches

Text #1	Text #2	Index	Words	Similarity?
<b>B04</b>	AD04	0,87	52	Yes
<b>B06</b>	AD06	0,69	130	Yes
<b>B08</b>	AD08	0,75	117	Yes
<b>B10</b>	AD10	0,69	78	Yes
<b>B12</b>	AD12	0,78	60	Yes
<b>B14</b>	AD14	0,69	92	Yes
<b>B16</b>	AD16	0,52	108	Yes
<b>B18</b>	AD18	0,76	44	Yes
<b>B20</b>	AD20	0,78	46	Yes
<b>B22</b>	AD22	0,71	36	Yes
<b>B24</b>	AD24	0,76	101	Yes
<b>B26</b>	AD26	0,81	121	Yes
<b>B28</b>	AD28	0,76	165	Yes
<b>B30</b>	AD30	0,59	75	Yes
<b>B32</b>	AD32	0,75	77	Yes
<b>B34</b>	AD34	0,76	101	Yes
<b>B36</b>	AD36	0,83	48	Yes
<b>B38</b>	AD38	0,75	56	Yes
<b>B40</b>	AD40	0,45	70	Yes
<b>B42</b>	AD42	0,54	56	Yes
<b>B44</b>	AD44	0,28	65	No

When both the press release and the news agency dispatch were obtained, they were compared and the results are shown in Image 7. The similarity index

between the news brief and agency dispatch is greater than the index between the news brief and press release in all cases but one.

Image 7 Jaccard binary index when comparing news briefs with agency dispatches (black columns) and with press releases (red columns)



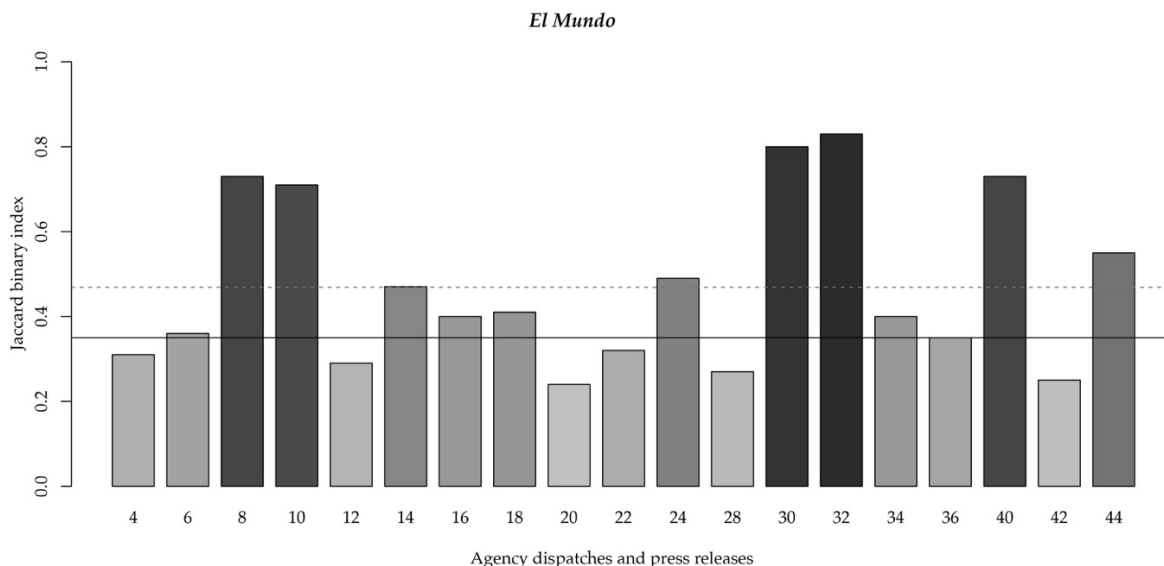
Source: Own elaboration

In the case of the newspaper *El Mundo*, the bylines carried by all but one of the news briefs analysed in this research were indicated as news agencies: EFE and in particular Europa Press.

The similarity algorithm was used to study the relationship between the agency dispatches feeding the

*El Mundo* news briefs and the relevant press releases. The results appear in Image 8 and Table VII. The similarity value is greater than 0.35 in 13 of the 19 pairs of compared texts, which represents 68.42% of the total. The similarity index average stands at 0.47, which is again higher than 0.35.

Image 8 Jaccard binary index when comparing the press releases with the agency dispatches



Source: Own elaboration

Table VII. Jaccard binary index, result of comparing the indicated press releases and the agency dispatches

Text #1	Text #2	Index	Words	Similarity?
AD04	PR04	0,31	681	No
AD06	PR06	0,36	276	Yes
AD08	PR08	0,73	166	Yes

AD10	PR10	0,71	345	Yes
AD12	PR12	0,29	165	No
AD14	PR14	0,47	615	Yes
AD16	PR16	0,40	484	Yes
AD16	PR18	0,41	339	Yes
AD20	PR20	0,24	266	No
AD22	PR22	0,32	276	No
AD24	PR24	0,49	458	Yes
AD28	PR28	0,27	305	No
AD30	PR30	0,80	324	Yes
AD32	PR32	0,83	162	Yes
AD34	PR34	0,40	248	Yes
AD36	PR36	0,35	178	Yes
AD40	PR40	0,73	185	Yes
AD42	PR42	0,25	207	No
AD44	PR44	0,55	223	Yes

**4. Conclusions**

Next, we will check if there are significant differences between both media in terms of similarity indexes. Table VIII includes a summary

indicating the arithmetic mean and the standard deviation for the similarity indexes between the news briefs and press releases (PR) and between the news briefs and agency news releases (AD) in both newspapers.

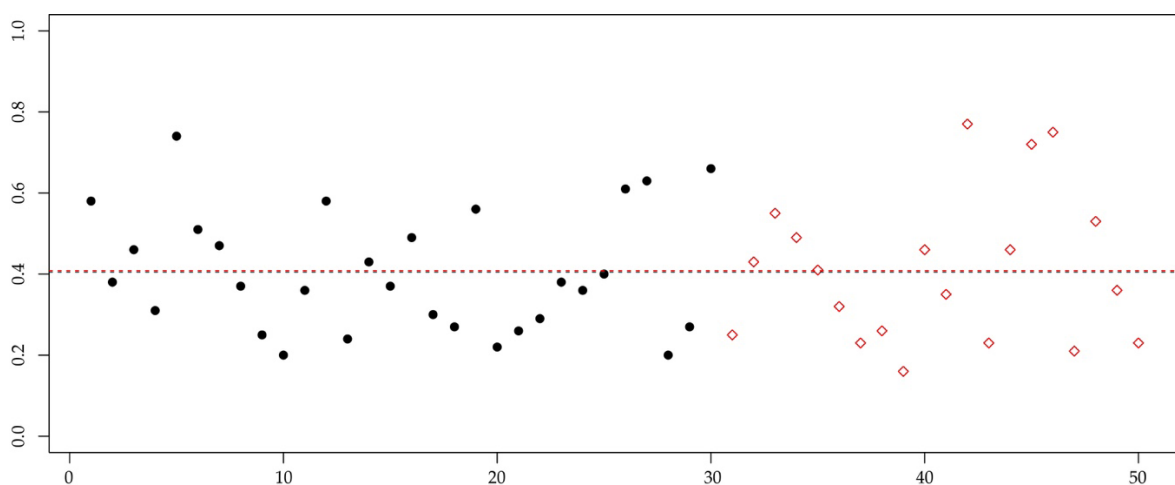
**Table VIII.** Summary of the most common statistics (mean and standard deviation) for similarity indices between news brief and press release and between news brief and agency dispatch, for *La Vanguardia* and *El Mundo*

	<i>La Vanguardia</i>		<i>El Mundo</i>	
	Average	St. Dev.	Average	St. Dev.
Press release	0,41	0,15	0,41	0,18
Ag. dispatch	0,46	0,19	0,69	0,14

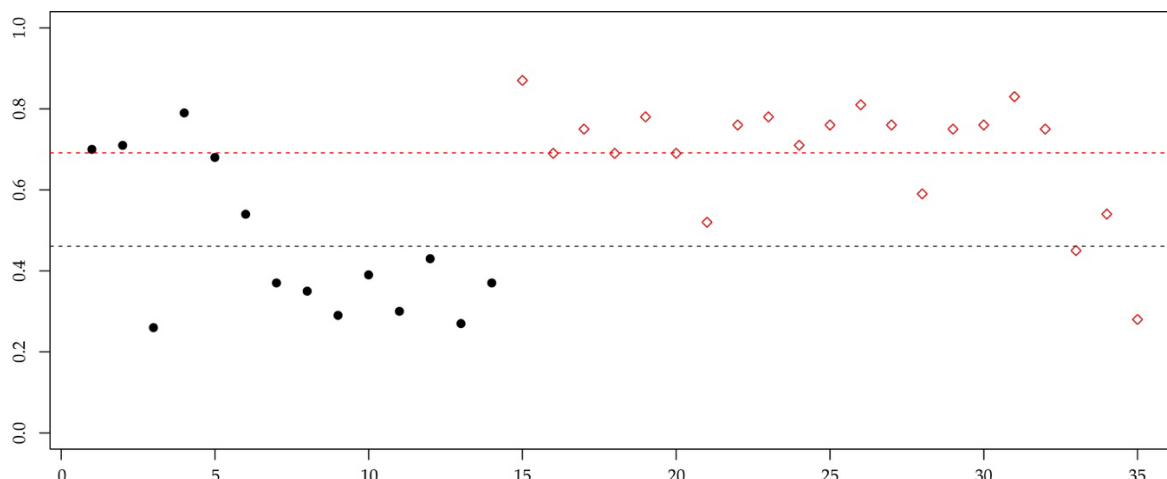
The similarity indices can be viewed separately in Images 9 and 10, which, respectively, were de-

signed for the press releases and agency dispatches.

**Image 9.** Similarity indices between news brief and press release in *La Vanguardia* (black dots) and *El Mundo* (red diamonds). The dashed black and red lines (effectively, superimposed over each other) represent the average values in each case



**Image 10** Similarity indices between news brief and agency dispatch in *La Vanguardia* (black points) and *El Mundo* (red diamonds). The dashed black and red lines represent the average values in each case



These visual representations show that the similarity indices between the news briefs and the press releases of *La Vanguardia* and *El Mundo* are close, although in *El Mundo* the dispersion of the data is somewhat greater. There is a noticeable difference in the similarity indices between the news briefs and the agency dispatches. The average value is 0.46 in *La Vanguardia*, while the value increases to 0.69 in *El Mundo*. A hypothesis test will be carried out to confirm or discard the insight derived from this difference.

Testing the hypothesis allows us to rigorously establish whether—according to scientific criteria—there are significant differences between *La Vanguardia* and *El Mundo* in regard to their respective similarity indices between their news briefs and the agency news releases. Our hypothesis testing relied on the so-called Welch-Satterthwaite test, which is summarized in the work of Mújica et al. (2014). The null hypothesis (H0) was that no differences existed in regard to the distribution of both sets, while the alternative hypothesis (H1) established that both sets were distributed differently. To test the hypothesis, a significance level of 10% was established and the free program R was used.

For the abovementioned data set, the  $p$ -value of the test is equal to 0.069%, which is clearly less than the 10% level of significance, meaning that the differences between both data sets are statistically and significantly different, such that it can be inferred that the similarities between news briefs and agency dispatches are greater in *El Mundo* than in *La Vanguardia*.

In 20 of the 52 studied news briefs, the byline given is “Editorial”; 35 of the 52 news briefs reproduce agency dispatches; 32 of these news briefs do attribute the source to the agencies (*El Mundo* surpasses *La Vanguardia* in this aspect: only one of the 22 studied news briefs carried the byline Editorial Board). Thus, most of the bylines in the news brief are agency (32). In these cases, the

news agency has mostly copied the original press release. Twenty-one of these 32 news briefs exhibit similarity with the corresponding press releases, which were presumably prepared by the news agencies (Europa Press, mostly). In the 20 cases that remain out of the 52 news briefs, the bylines of the news briefs correspond to the Editorial board.

In 14 of the 19 news briefs giving the byline to the editors in the *La Vanguardia* Economy section, there is similarity with the original press release. What is more, 3 of these 14 news briefs are actually from the agency. As far as *El Mundo* is concerned, only 1 of the studied news briefs gave the byline to the editorial board, although this piece presents striking similarities to the original press release.

The interviewees coincide in their interpretation of these data. Those who work in the sector have become familiar with this *modus operandi*, and hardly anyone is surprised anymore that the media industry works in this way. The mechanics of this process make it fast and cheap: It requires little manpower, which in addition does not require great training and skill to function in this manner.

As has happened historically (McCombs and Shaw, 1972, Manning, 2001), companies insist that their press offices and communications departments guarantee a certain presence in the media. For this reason, these professionals have learned to present their information in a way that leads journalists to believe the material needs no rewording, as the interviewees recall. They further add that the agencies are in a hurry to be the first to distribute any news, which presumably tempts them to copy and paste the press releases. As the participants in this investigation emphasize, journalists follow similar behaviour when these dispatches reach their newsrooms. Thus, the interviewees conclude that the press release—despite having passed through two newsrooms (the agency and the intermediary)—undergoes virtually no alteration.

## 5. Referencias bibliográficas

- Bacon, Wendy et al. (15 de marzo del 2010). Spinning the Media: Key findings in a week in the life of the media. Recuperado de: <https://www.crikey.com.au/2010/03/15/spinning-the-media-key-findings-in-a-week-in-the-life-of-the-media/>
- Barrón-Cedeño, Alberto et al. (2013). «Methods for cross-language plagiarism detection». *Knowledge-Based Systems*, 50, 211-217. <https://www.sciencedirect.com/science/article/abs/pii/S0950705113002001>
- Busquet, Jordi; Medina, Alfons y Sort, Josep. (2006). *La recerca en comunicació. Què hem de saber? Quins passos hem de seguir?* (Vol. 95). Barcelona: Editorial UOC.
- Csikszentmihalyi, Mihaly y Larson, Reed. (2014). «Validity and reliability of the experience-sampling method». En: Csikszentmihalyi, Mihaly. *Flow and the foundations of positive psychology*. Dordrecht (Holanda): Springer, pp. 35-54.
- Chacón, Inmaculada. (1998). «La fiabilidad del proceso documental del discurso periodístico como fuente de información». (Tesis doctoral). Universidad Complutense de Madrid, España. Recuperado de: <https://eprints.ucm.es/1796/>
- De Fontcuberta, Mar. (1993, 2000). *La noticia. Pista para percibir el mundo*. Barcelona: Paidós.
- Durandin, Guy. (1982). *La mentira en la propaganda política y en la publicidad*. Barcelona: Paidós.
- Estudio General de Medios. Resumen general. Febrero a noviembre 2017. Madrid: Asociación para la investigación de medios de comunicación (AIMC).
- Fonnegra, Nazly María. (23 de febrero del 2010). Lead y estructura de la noticia. *Nazly María Fonnegra*. Recuperado de: <http://nazlymariafonnegra.blogspot.com/2010/02/lead-y-estructura-de-la-noticia.html>
- García Márquez, Gabriel. (2010). *Yo no vengo a decir un discurso*. Barcelona: DeBolsillo.
- Gomis, Llorenç. (2008). *Teoría de los géneros periodísticos*. Barcelona: Editorial UOC.
- Grijelmo, Álex. (2006). *El estilo del periodista*. Madrid: Taurus.
- Jaraba, Gabriel. (2009). Estructura de la noticia: el lead y el cuerpo. *Periodismo UAB wiki*. Recuperado de: <http://periodismouab.wikispaces.com/space/content?o=20>
- Manning, Paul. (2001). *News and news sources. A critical introduction*. Londres: Sage.
- McCombs, Maxwell E. y Shaw, Donald L. (1972). «The agenda-setting function of mass media». *Public Opinion Quarterly*, 36 (2), 176-187.
- McChesney, Robert. (1999). *Rich media, poor democracy: Communication politics in dubious times*. Chicago: University of Illinois Press.
- Mújica, Luis E., et al. (2013). «A structural damage detection indicator based on principal component analysis and statistical hypothesis testing». *Smart materials and structures*, 23 (2), 025014.
- Niwattanakul, Suphakit et al. (2013). «Using of Jaccard Coefficient for Keywords Similarity». En: *Proceedings of the International MultiConference on Engineers and Computer Scientists*. China: Hong Kong.
- Parratt, Sonia. (2008). *Géneros periodísticos en prensa*. Quito: Ediciones Ciespal.
- Pérez, Concha. (2005). «Estrategias de selección y tratamiento de las fuentes en las secciones periodísticas de *El Mundo* y *El País*». (Tesis doctoral). Universidad de Sevilla, España. Recuperado de: <https://idus.us.es/xmlui/handle/11441/24400>
- Romero, Luis Miguel. (2014). «Pragmática de la desinformación: Estratagemas e incidencia de la calidad informativa de los medios». (Tesis doctoral). Universidad de Huelva, España. Recuperado de: <http://rabida.uhu.es/dspace/handle/10272/9605>
- Sancha, David. (2005). «El uso de la información de agencia en las ediciones electrónicas de diarios en España: estudio comparativo de las páginas web de *El Periódico*, *El Mundo* y *La Vanguardia*». *Athenea digital: revista de pensamiento e investigación social*, (8), 234-267. <https://pdfs.semanticscholar.org/ad80/f42993827a815a8afab84dba5e-a4ef8257c4.pdf>
- Sánchez-Vega, Fernando et al. (2013). «Determining and characterizing the reused text for plagiarism detection». *Expert Systems with Applications*, 40 (5), 1804-1813.
- Scholz, Trebor. (2013). *Digital labor: The Internet as playground and factory*. Nueva York: Routledge.
- Schwoebel, Jean. (1971). *La prensa, el poder y el dinero*. Barcelona: Dopesa.
- Starr, Paul. (2009). «Adiós a la era de los periódicos (bienvenida una nueva era de corrupción)». En: Espada, Arcadi y Hernández-Busto, Ernesto. (eds.). *El fin de los periódicos. Crisis y retos del periodismo actual*. Barcelona: Duomo Ediciones.
- Tellería Roca, Evelio. (1986). *Diccionario periodístico*. Santiago de Cuba: Editorial Oriente.
- Varios autores (2004). *Libro de redacción de La Vanguardia*. Barcelona: Ariel.
- Varios autores. (2002). *Libro de estilo de El Mundo*. Sevilla: Comunicación y Proyectos Editoriales en Contexto.
- Wang, Shuai et al. (2013). Combination of VSM and Jaccard coefficient for external plagiarism detection. En: *IEEE International Conference on Machine Learning and Cybernetics*, Vol. 4, pp. 1880-1885.
- Weill, Georges. (2007). *El periódico. Orígenes, evolución y función de la prensa periódica*. Sevilla: Comunicación Social.
- Wimmer, Roger y Dominick, Joseph. (2013). *Mass media research: An introduction*. Boston: Cengage Learning.
- Wolf, Mauro. (1987). *La investigación de la comunicación de masas: crítica y perspectivas*. Barcelona: Paidós.